PLUG-IN ERROR BOUNDS FOR A MIXING DENSITY ESTIMATE IN R^d, AND FOR ITS DERIVATIVES

Yannis G. Yatracos

Faculty of Communication and Media Studies

Cyprus University of Technology

October 26, 2015

Summary

A mixture density, f_p, is estimable in $R^d,\ d \geq 1$, but an estimate for the mixing density, p, is usually obtained only when d is unity; h is the mixture's kernel. When f_p's estimate has form $f_{\hat{p}_n}$ and p is $\tilde{q}$-smooth, vanishing outside a compact in R^d, plug-in upper bounds are obtained herein for the L_u-error (and risk) of $\hat{p}_n$ and its derivatives; $d \geq 1, 1 \leq u \leq \infty$. The bounds depend on $f_{\hat{p}_n}$'s L_u-error (or risk), h's Fourier transform, $\tilde{h}$, and the bandwidth of kernel K used in approximations. The choice of $\hat{p}_n$, via $f_{\hat{p}_n}$, suggests that $\hat{p}_n$'s error rate could be only nearly optimal when $f_{\hat{p}_n}$ is optimal, but competing estimates and their error rates may not be available for $d > 1$. In examples with d unity, the upper bound is optimal when h is super smooth, misses the optimal rate by the factor $(\log n)^{\xi},\ \xi > 0$, when h is smooth, and is satisfactory when $\tilde{h}$ has periodic zeros.

1 Introduction

In the mixing problem, random vectors Y and X have densities, respectively, p and f_p such that

$$f_p(x) = \int_{R^d} h(x|y)p(y)dy, \ p \in \mathcal{P}, \ d \geq 1. \tag{1}$$

Independent copies $X_1, \ldots, X_n$ of X are observed and the goal is to estimate p, its s-th mixed partial derivative, $p^{(s)}$, and calculate the estimation errors. Usually h is assumed known, with non-vanishing Fourier transform $\tilde{h}$. The classic approach in the 1-dimensional deconvolution is to obtain a kernel estimate for p by assuming, in addition, that

$$X = Y + \epsilon; \tag{2}$$

Y and ϵ are independent and h is the density of the error ϵ, such that $h(x|y)$ has form $h(x-y)$ in (1). Robbins (1948, page 366) provided an example showing that Y, ϵ need not be independent in (2) for $h(x|y)$ to have form $h(x-y)$.

Research has been devoted mainly to the 1-dimensional problem. However, it is crystal clear that the X-observations can be used to estimate f_p in higher dimension, e.g., using maximum likelihood or minimum distance methods. If an estimate with form $f_{\hat{p}_n}$ is obtained, then $\hat{p}_n$ estimates p and the problem that has not been tackled so far in the literature is to derive "plug-in" L_u-$||\cdot||_u$-rates of convergence for $\hat{p}_n$ and $\hat{p}_n^{(s)}$ in R^d from those of $f_{\hat{p}_n}, 1 \leq u \leq \infty$.

This problem is addressed herein when p is $\tilde{q}$-smooth, vanishing outside a compact in R^d, and assuming as in Zhang (1990) that $h(x|y) = h(x-y)$. An upper bound for $||\hat{p}_n - p||_u$ is provided that depends on $||f_{\hat{p}_n} - f_p||_u$, $\tilde{h}$ and a smoothing parameter b_n of kernel K used in approximations; b_n is chosen to obtain the best rate. Bounds follow for $||\hat{p}_n^{(s)} - p^{(s)}||_u$ in probability and for its expected value (i.e., in risk). The bounds hold also in L_∞-distance for any $\tilde{q}$-smooth density p in R^d vanishing at infinities. The use of K and its Fourier transform, $\tilde{K}$, allows to connect the error rate of $\hat{p}_n$ with that of $f_{\hat{p}_n}$.

The class of estimates $\hat{p}_n$ obtained via $f_{\hat{p}_n}$ is a subset of all p's estimates, thus the fastest rate of convergence within this class may be larger than the minimax rate, available so far *only* when d is unity, i.e., pointwise (Carrol and Hall, 1988) and in L_u-risk (Fan, 1991,

1992, 1993). However, if $f_{\hat{p}_n}$ is optimal, the difference of $\hat{p}_n$'s error from the optimal is not expected to be substantial. For example, if h is super-smooth and $f_{\hat{p}_{n,\delta}}$ converges to f_p with rate $n^{-\delta}(\log n)^{\zeta}$, in probability or in risk, then all $\hat{p}_{n,\delta}$ have error rate proportional to $(\log n)^{-\tilde{q}}$ for any δ, except for a constant factor independent of n; this rate is optimal compared with the lower L_u-error rate when d is unity. This is confirmed in L_1-distance for the sieves MLE and the generalized MLE, $f_{\hat{p}_n}$, of 1-dimensional Gaussian mixture (Genovese and Wasserman, 2000, Ghosal and van der Vaart, 2001, and Zhang, 2009). If h is smooth and d is unity, the rates differ from the optimal minimax by the factor $(\log n)^{\xi}$, $\xi > 0$. When $\tilde{h}$ has finite number of zeros in every compact in R^d, a general error bound for $\hat{p}_n$ is provided. In each particular problem, this bound will depend on $f_{\hat{p}_n}$'s error and the implementation. When h is smooth and $\tilde{h}$'s zeros are periodic in R, an upper error bound is obtained with an implementation herein.

Robbins (1955, 1964) introduced initially the 1-dimensional mixing density problem but later used in (1) cumulative distribution functions and obtained a minimum distance estimate for the mixing distribution. For the mixing density and the deconvolution problems, consistent estimates have been provided and, when p is $\tilde{q}$-smooth, optimality of the error rates has been established for smooth and super-smooth h, pointwise and in L_u-distance, among others by Carroll and Hall (1988), Devroye (1989), Stefanski and Carroll (1990), Zhang (1990), Fan (1991, 1992), Hesse (1995) and Loh and Zhang (1996, 1997), $1 \leq u \leq \infty$. Devroye (1989) showed in particular that one can construct a consistent kernel estimate of p when the set $\{t : \tilde{h}(t) = 0\}$ has Lebesgue measure zero. For finite mixture models, Chen (1995) provided an optimal minimum distance estimate for p's cumulative distribution function. More recent work using kernel estimates includes, among others, Delaigle and Gijbels (2002) and Meister (2006). Johannes (2009) estimated non-parametrically p when ϵ's distribution in (2) is estimated. Comte and Lacour (2013) and Rebelles (2015) study deconvolution in R^d when p lives in an anisotropic Nikolskii class that makes Y's coordinates independent, p product of densities and the obtained error rates those for 1-dimensional deconvolution; $d \geq 1$.

Groneboom and Jongbloed (2003) provide "a type of kernel density estimates" with rates of convergence pointwise when p has two derivatives and h is uniform. Hall and

Meister (2007) present a new estimate for p using *ridging*, "not involving kernels in any way", used also when $\tilde{h}$ has periodic zeros. Meister (2008) proposes also an estimate for p using local polynomials when $\tilde{h}$ has periodic zeros. Under additional assumptions on either p or h, the estimates in Hall and Meister (2007, see page 1542, lines -3, -2) and in Meister (2008, see the Introduction) are optimal but the assumptions and the rates are different. When $\tilde{h}$ has periodic zeros, our bounds may be improved using the additional assumptions and the implementations in Hall and Meister (2007) and in Meister (2008) but the price will be the restrictions on either the class $\mathcal{P}$ or on h; see also Remark 4.2 for additional reasons we did not follow this path.

Notations, definitions and tools appear in section 2. Upper bounds are provided in section 3 for non-vanishing $\tilde{h}$ and in section 4 when $\tilde{h}$ has zeros. Proofs and auxiliary results are in the Appendix.

2 Notation, Definitions, the Tools

All functions are assumed to be measurable and when the domain of integration is R^d it is omitted. All densities are defined with respect to Lebesgue measure. For any function g its Fourier transform is $\tilde{g}$. The vectors $X,\ Y$ take values, respectively in $\mathcal{X},\ \mathcal{Y}$, both sets in R^d. $C,\ c$ denote generic constants, and for positive $a, b,\ a \sim b$ means $C_1 b \leq a \leq C_2 b$, with C_1, C_2 positive fixed constants. When ρ is a distance, the expression $\rho(\hat{p}_n, p)$ *is bounded by* a_n *in probability and in risk* means, respectively,

$$\lim_{n\to\infty} P[\rho(\hat{p}_n, p) > Ca_n] = 0, \qquad E\rho(\hat{p}_n, p) \leq Ca_n; \tag{3}$$

$a_n > 0,\ E$ denotes expected value.

Definition 2.1 For densities p_1, p_2 defined in $\mathcal{Y}(\subset R^d)$ their L_u-distance is

$$||p_1 - p_2||_u = [\int_{\mathcal{Y}} |p_1(w) - p_2(w)|^u dw]^{1/u},\ 1 \leq u < \infty. \tag{4}$$

The L_∞- distance is

$$||p_1 - p_2||_\infty = \sup_{w\in\mathcal{Y}} |p_1(w) - p_2(w)|. \tag{5}$$

The Hellinger distance is

$$H(p_1,p_2)=[\int_{\mathcal{Y}}(\sqrt{p_1(y)}-\sqrt{p_2(y)})^2dy]^{1/2}, \tag{6}$$

and

$$||p_1-p_2||_1 \leq 2H(p_1,p_2). \tag{7}$$

For integrable, real valued functions $h,\ p$ defined in R^d, their convolution

$$h*p(x)=\int h(x-y)p(y)dy.$$

For the support of $h*p$ it holds

$$support(h*p)\subset \overline{\{x\in R^d: x=y+u,\ y\in support(p),\ u\in support(h)\}}; \tag{8}$$

$\overline{A}$ denotes the closure of A. From (8) it follows that only when the supports of h and p are both bounded then $h*p$'s support is bounded. Also that when the support of h includes the value zero, then p's support is a subset of $h*p$'s support.

If $x=(x_1,\ldots,x_d)\in R^d,\ a\in R$ and $s=(s_1,\ldots,s_d)$ is a d-tuple of non-negative integers,

$$x^s=(x_1^{s_1},\ldots,x_d^{s_d}),\ xs=x_1s_1+\ldots+x_ds_d,\ ax=(ax_1,\ldots,ax_d),\ [s]=s_1+\ldots+s_d.$$

For a real valued function g defined in R^d let $g^{(s)}(x_0)$ denote the s-th order mixed partial derivative of g at x_0, i.e.

$$g^{(s)}(x_0)=\frac{\partial^{[s]}g(x_0)}{\partial x_1^{s_1}\ldots x_d^{s_d}}.$$

Let $K(x)$ be a symmetric function defined in R^d at least q times continuously differentiable with bounded Fourier transform $\tilde{K}$ having compact support $[-M,M]^d,\ M>0$, such that for $s\in(R^+)^d$,

$$\int K(x)dx=1, \quad \int x^sK(x)dx=0,\ [s]=1,\ldots,q, \int(|x|^q+|x|^{q+1})K(x)dx<\infty. \tag{9}$$

Kernel K can be obtained by taking d-fold products of Devroye's trapezoidal kernel (Devroye, 1992) and making smooth enough the linear leg of the trapezoid (Devroye, 2013). For any positive number b_n, let

$$K_n(x)=b_n^{-d}K(xb_n^{-1}), \tag{10}$$

with b_n decreasing to 0 as n increases.

3 Upper rates of convergence, $\tilde{h} \neq 0$

Let $X_1, \ldots, X_n$ be independent, identically distributed observations with values in $\mathcal{X}(\subset R^d)$, $d \geq 1$ and density f_p satisfying (1) with p defined on $\mathcal{Y}(\subset R^d)$. It is not assumed that (2) holds but instead, as in Robbins (1964) and Zhang (1990), that $h(x|y)$ is a location family with location y.

The Assumptions:

$(\mathcal{A}1)$ $\tilde{h} \neq 0$, $||\tilde{h}||_2 < \infty$ and $h(x|y) = h(x-y)$, thus (1) becomes

$$f_p(x) = \int_{\mathcal{Y}} h(x-y)p(y)dy = h * p(x). \tag{11}$$

$(\mathcal{A}2)$ p has all s-th mixed order partial derivatives for $0 \leq [s] \leq q$, with the q-th mixed order having modulus of continuity w_q.

$(\mathcal{A}3)$ $\mathcal{Y}$ is compact,

$(\mathcal{A}4)$ $\mathcal{Y} \subset \mathcal{X}$.

$(\mathcal{A}5)$ $f_{\hat{p}_n}$ is an estimate of f_p such that either $||f_{\hat{p}_n} - f_p||_u \sim a_n$ in probability, or $E||f_{\hat{p}_n} - f_p||_u \sim a_n$, with a_n converging to zero as n increases, $1 \leq u \leq \infty$.

Assumptions $(\mathcal{A}1) - (\mathcal{A}3)$ are used in deconvolution problems for which $(\mathcal{A}4)$ usually holds since the error ϵ in (2) takes also the value zero. In $(\mathcal{A}5)$, $f_{\hat{p}_n}$ can be, e.g., a minimum distance estimate. Identifiability of p follows from $(\mathcal{A}1)$.

Let $\tilde{h}$ and $\tilde{K}_n$ be, respectively, the Fourier transforms of h and K_n (see (10)). Since $\tilde{h} \neq 0$, let ψ_n be the inverse Fourier transform of

$$\tilde{\psi}_n = \frac{\tilde{K}_n}{\tilde{h}}. \tag{12}$$

By the convolution theorem,

$$\psi_n * h = K_n. \tag{13}$$

An upper bound for $||\psi_n||_1$ is obtained. The set $[-M, M]^d$ is the support of $\tilde{K}$.

Lemma 3.1 *Under* $(\mathcal{A}1)$,

$$||\psi_n||_1 \leq C \cdot [\int_{[-\frac{M}{b_n}, \frac{M}{b_n}]^d} |\tilde{K}(tb_n)|^2 |\tilde{h}(t)|^{-2} dt\ ]^{1/2} \leq C \cdot \frac{\sup_{t \in [-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1}}{b_n^{.5d}}. \tag{14}$$

In the next proposition, the general bound for $||\hat{p}_n - p||_u$ is provided when $\tilde{h} \neq 0$. The quality of $f_{\hat{p}_n}$ will reflect on the quality of $\hat{p}_n$. If $f_{\hat{p}_n}$ is optimal for estimating f_p, then $\hat{p}_n$ is the best one can do within this class of estimates without additional effort.

Proposition 3.1 *a) Under assumptions* $(\mathcal{A}1) - (\mathcal{A}5)$,

$$||\hat{p}_n - p||_u \leq C[b_n^q w_q(b_n) + ||\tilde{\psi}_n^*||_2 \; a_n] \leq C[b_n^q w_q(b_n) + \frac{\sup_{t \in [-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1}}{b_n^{.5d}} ||f_{\hat{p}_n} - f_p||_u]; \tag{15}$$

$[-M, M]^d$ *is* $\tilde{K}$*'s support,* $1 \leq u \leq \infty$.

b) Under assumptions $(\mathcal{A}1), (\mathcal{A}2)$ *and* $(\mathcal{A}5)$, *with* p *defined in* R^d *and bounded, the upper bound (15) remains valid in* L_∞*-distance.*

Proposition 3.1 provides bounds on $||\hat{p}_n - p||_u$ in probability and on $E||\hat{p}_n - p||_u$ using, respectively, the bounds for $||f_{\hat{p}_n} - f_p||_u$ and for $E||f_{\hat{p}_n} - f_p||_u$. Careful choice of b_n determines the least upper bound (15). When $\tilde{h}(t)$ varies exponentially as t increases, the term with $\tilde{h}$ in (15) determines the upper bound. For algebraic variation of $\tilde{h}(t)$ as t increases, b_n satisfies

$$b_n^q w_q(b_n) \sim \frac{\sup_{t \in [-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1}}{b_n^{.5d}} a_n. \tag{16}$$

The obtained convergence rates of the error and risk are satisfactory for super-smooth and smooth h.

Upper bounds for $||\hat{p}_n - p||_u$ in probability and for $E||\hat{p}_n - p||_u$ are now given explicitly as function of the bound a_n in $(\mathcal{A}5)$ for super-smooth and smooth h.

Non-Oscillatory Smooth and Super-smooth Models $(\mathcal{M}1), (\mathcal{M}2)$

The terms "non-oscillatory" and "oscillatory" models are introduced in Hall and Meister (2007) but we make a model modification. Let $0 < C_1 \leq C_2 < \infty$, $|t| = (|t_1, \ldots, |t_d|)$, $k > 0$, $\alpha_j \geq 0$, $\beta_j > .5, j = 1, \ldots, d$, $d\bar{\alpha} = \sum_{j=1}^d \alpha_j$, $d\bar{\beta} = \sum_{j=1}^d \beta_j$.

$(\mathcal{M}1)$ h is super-smooth when $\tilde{h} \neq 0$ and for large $|t|$-values, $d\bar{\alpha} > 0$,

$$C_1 e^{-\sum_{j=1}^d \alpha_j |t_j|^k} \Pi_{j=1}^d |t_j|^{\beta_j} \leq |\tilde{h}(t_1, \ldots, t_d)| \leq C_2 e^{-\sum_{j=1}^d \alpha_j |t_j|^k} \Pi_{j=1}^d |t_j|^{\beta_j}. \tag{17}$$

$(\mathcal{M}2)$ h is smooth when $\tilde{h} \neq 0$ and for large $|t|$-values

$$C_1 \Pi_{j=1}^d |t_j|^{-\beta_j} \leq |\tilde{h}(t_1, \ldots, t_d)| \leq C_2 \Pi_{j=1}^d |t_j|^{-\beta_j}, \tag{18}$$

Proposition 3.2 *a) Assume that* $(\mathcal{A}1) - (\mathcal{A}5)$ *hold.*

i) For super-smooth h *from model* $(\mathcal{M}1)$, *an upper bound in probability on* $\hat{p}_n$*'s estimation error is*

$$||\hat{p}_n - p||_u \leq C_{\bar{\alpha},d,k,M} \cdot (\log a_n^{-1})^{-q/k} w_q[C(\log a_n^{-1})^{-1/k}]. \tag{19}$$

When $w_q(b_n) \sim b_n^\gamma, 0 < \gamma < 1, \tilde{q} = q + \gamma,$

$$||\hat{p}_n - p||_u \leq C_{\bar{\alpha},d,k,M} \cdot (\log a_n^{-1})^{-\tilde{q}/k}. \tag{20}$$

The dimension d *affects only constant* $C_{\bar{\alpha},d,k,M},\ d\bar{\alpha} = \sum_{j=1}^d \alpha_j.$

ii) For smooth h *from model* $(\mathcal{M}2)$, *an upper bound on* $||\hat{p}_n - p||_u$ *is obtained when* b_n *satisfies*

$$b_n^q w_q(b_n) \sim \frac{a_n}{b_n^{d\bar{\beta}+.5d}},\ d\bar{\beta} = \sum_{j=1}^d \beta_j.$$

When $w_q(b_n) = b_n^\gamma,\ 0 < \gamma < 1,$ *an upper bound in probability on* $\hat{p}_n$*'s error is*

$$||\hat{p}_n - p||_u \leq c_M a_n^{\tilde{q}/(\tilde{q}+d\bar{\beta}+.5d)},\ \tilde{q} = q + \gamma. \tag{21}$$

iii) When $E||f_{\hat{p}_n} - f_p||_u \leq a_n$ *and* $w_q(b_n) \sim b_n^\gamma$, *then the bounds in (20) and (21) hold also for* $E||\hat{p}_n - p||_u$.

All the bounds in i)-iii) are valid for $1 \leq u \leq \infty$.

b) Under assumptions $(\mathcal{A}1), (\mathcal{A}2)$ *and* $(\mathcal{A}5)$, *with* p *defined in* R^d *and bounded, the upper bounds in a) remain valid in* L_∞*-distance.*

Remark 3.1 *Model* $(\mathcal{M}1)$ *can be enlarged, with* k *in (17) replaced by positive* $k_j, j = 1, \ldots, d$. *Then, upper bounds (19), (20) remain valid with* $\max\{k_1, \ldots, k_d\}$ *replacing* k. *In the proof,* k *of the upper bound in (50) will be replaced by* $\max\{k_1, \ldots, k_d\}$.

Bounds on $\hat{p}_n$'s error follow.

Corollary 3.1 *a) Assume* $(\mathcal{A}1)-(\mathcal{A}5)$ *hold,* $w_q(b)=b^\gamma, 0<\gamma<1$, $\tilde{q}=q+\gamma$, $s=(s_1,\ldots,s_d)$ *is a d-tuple of non-negative integers,* $[s]=s_1+\ldots+s_d\leq q$.
i) If $||\hat{p}_n-p||_u\leq\delta_n$ *in probability, then*

$$||\hat{p}_n^{(s)}-p^{(s)}||_u\leq C\cdot\delta_n^{\frac{\tilde{q}-[s]}{\tilde{q}}} \tag{22}$$

in probability.
ii) If $E||\hat{p}_n-p||_u\leq\delta_n$, *then*

$$E||\hat{p}_n^{(s)}-p^{(s)}||_u\leq C\cdot\delta_n^{\frac{\tilde{q}-[s]}{\tilde{q}}}. \tag{23}$$

b) Under assumptions $(\mathcal{A}1)-(\mathcal{A}3)$ *and* $(\mathcal{A}5)$, *with* p *defined in* R^d, *bounded and with its derivatives vanishing at infinities, the upper bounds in a) remain valid in* L_∞*-distance.*

The next result indicates the reason that in R, estimates of p and $p^{(s)}$ are frequently minimax optimal when h is super-smooth.

Corollary 3.2 *Under the assumptions in Proposition 3.2 a) i) and Corollary 3.1 and if* $||\hat{f}_n-f_p||_u\sim n^{-\delta}$ *in probability,* $0<\delta<1$,

$$||\hat{p}_n^{(s)}-p^{(s)}||_u\leq C_{\bar{\alpha},d,k,M}(\delta\log n)^{-(\tilde{q}-[s])/k},\ [s]\geq 0. \tag{24}$$

If $E||\hat{f}_n-f_p||_u\sim n^{-\delta}$, *the upper bound in (24) is valid for the risk* $E||\hat{p}_n^{(s)}-p^{(s)}||_u$.
When $d=1$, $\hat{p}_n^{(s)}$ *is minimax optimal for any* $\delta>0$.

Existing estimates $f_{\hat{p}_n}$ in the literature and their Hellinger upper error rates provide below L_1-upper error rates for $\hat{p}_n$.

Example 3.1 *When* p *is defined on a compact subset of the real line, sieve maximum likelihood estimates for location-scale Gaussian mixtures, with known lower bound of scale, are obtained by Genovese and Wasserman (2000) with Hellinger upper error rates in probability* $(\log n)^{1/4}/n^{1/4}$. *The rates are improved by Ghosal and van der Vaart (2001) using*

generalized maximum likelihood estimate to $(\log n)^{\kappa}/\sqrt{n},\ \kappa \geq 1$*. Zhang (2009) improved the rates when the* m*-th weak moment of* p *is bounded, i.e.,*

$$x^m \int_x^{\infty} p(x)dx = O(1), \tag{25}$$

obtaining upper rate $(\log n)^{\kappa_m'}/n^{m/2(1+m)}, \kappa_m' = (2+3m)/(4+4m)$*. The presented rates in Hellinger distance are* $(\log n)^{\zeta}/n^{\delta},\ 0 < \delta \leq .5,\ \zeta > 0$*, and from (7), these bounds hold also for* L_1*-distance. Thus, from (24) for* $\tilde{q}$*-smooth density* p *and* $p^{(s)}$ *the upper rates for the corresponding estimates are, respectively,* $(\log n)^{-\tilde{q}/2}$ *and* $(\log n)^{-(\tilde{q}-[s])/2}$*, irrespectively of the particular* δ*-values.*

The error rates for f_p's estimates in Example 3.1 and additional results in the literature, e.g., Ibragimov (2001), suggest to use $a_n \sim n^{-1/2}(\log n)^{\zeta},\ 0 < \zeta$, to evaluate the obtained error bounds.

Example 3.2 *Assume that* $a_n \sim n^{-1/2}(\log n)^{\zeta}$ *in probability,* $d = 1, w_q(b) = b^{\gamma}, \gamma > 0, \tilde{q} = q + \gamma$*. Then:*

a) for h *the standard normal,* $\tilde{h}(t) \sim e^{-t^2}$ *for large* $|t|$*, and from (20), (22) in probability*

$$||\hat{p}_n^{(s)} - p^{(s)}||_1 \leq C(\log n)^{(\tilde{q}-[s])/2},\ [s] \geq 0.$$

b) *for* h *the Cauchy,* $\tilde{h}(t) \sim e^{-|t|}$ *for large* $|t|$*, and from (20), (22) in probability*

$$||\hat{p}_n^{(s)} - p^{(s)}||_1 \leq C(\log n)^{\tilde{q}-[s]},\ [s] \geq 0.$$

c) *for* h *the exponential,* $\tilde{h}(t) \sim |t|^{-\beta}$ *for large* $|t|$*, and from (21) in probability*

$$||\hat{p}_n^{(s)} - p^{(s)}||_1 \leq C\frac{(\log n)^{\xi}}{n^{(\tilde{q}-[s])/(2\tilde{q}+2\beta+1)}},\ \xi = \zeta(\tilde{q}-[s])/(\tilde{q}+\beta+.5),\ [s] \geq 0.$$

The bound in c) misses the minimax rates by the factor $(\log n)^{\xi}$.

The bounds in a)-c) remain valid when a_n *is the risk rate.*

4 Upper convergence rates when $\tilde{h}$ has zeros

Replace $(\mathcal{A}1)$ by

($\mathcal{A}1^*$) p is identifiable and $\tilde{h}$ has a finite number of roots-curves in any compact in R^d, at distance at least $\delta > 0$ in each compact, $||\tilde{h}||_2 < \infty,\ h(x|y) = h(x-y)$.

The requirement that roots-curves are δ-distant holds, e.g., for periodic roots in R. ($\mathcal{A}1^*$) will allow us to separate $\tilde{h}$ in 2 parts: one defined on a compact where the approach in the previous section will be used, and the parts in the tails that will be determined to have negligible effect on the upper bound of the error.

For M_n increasing to infinity let

$$\tilde{h}_n(t) = \tilde{h}(t)I(||t|| \leq M_n), \tag{26}$$

$$T(M_n) = ||\tilde{h} - \tilde{h}_n||_2 = [(\int_{[-\infty,-M_n]^d} + \int_{[M_n,\infty]^d})|\tilde{h}(y)|^2 dy]^{1/2}; \tag{27}$$

$||\cdot||$ is the sup-norm in R^d, I is the indicator function.

Denote by r_j a curve-variety of roots in $[-M_n, M_n]^d$ and let $v_{n,j}$ be positive numbers, $j = 1, \ldots, N(M_n)$, $v_n^* = (v_{n,1}, \ldots, v_{n,M_n})$ (abuse of notation, using v_n^* instead of v_n). For every roots-curve r_j, let R_j be the the region around it with $|\tilde{h}(t)|$ less than $v_{n,j}$,

$$R_j = \{t : |\tilde{h}(t)| \leq v_{n,j}, ||t - r_j|| = \inf_{\{k=1,\ldots,N(M_n)\}} ||t - r_k||\},\ j = 1, \ldots, N(M_n). \tag{28}$$

Define

$$\tilde{h}_n^*(t) = \tilde{h}_n(t)I(t \in [-M_n, M_n]^d - \cup_{j=1}^{N(M_n)} R_j) + \sum_{j=1}^{N(M_n)} v_{n,j} I(t \in R_j). \tag{29}$$

Thus, $\tilde{h}_n^*$ and $\tilde{h}_n$ differ on $R_j,\ j = 1, \ldots, N(M_n)$. Let

$$S(M_n) = ||\tilde{h}_n - \tilde{h}_n^*||_2 = [\sum_{j=1}^{N(M_n)} \int_{R_j} (v_{n,j} - |\tilde{h}(t)|)^2 dt]^{1/2}. \tag{30}$$

Let h_n^* be the inverse Fourier transform of $\tilde{h}_n^*$ and let $K,\ K_n$ be as previously defined, with $\tilde{K}$ vanishing outside $[-M, M]^d$. Let ψ_n^* be the inverse Fourier transform of

$$\tilde{\psi}_n^* = \tilde{K}_n / \tilde{h}_n^*. \tag{31}$$

From the convolution theorem,

$$K_n = \psi_n^* * h_n^*. \tag{32}$$

Since $\tilde{h}_n^*$ vanishes outside $[-M_n, M_n]^d$ and $\tilde{K}_n(x)$ equals $\tilde{K}(xb_n)$, for (31) to hold,

$$M_n = \frac{2M}{b_n^{2m}} \geq \frac{M}{b_n}, \; m \geq .5; \tag{33}$$

m to be determined.

Proposition 4.1 *Under the assumptions* $(\mathcal{A}1^*)$ *and* $(\mathcal{A}2) - (\mathcal{A}5)$, *with* $a_n \sim ||f_{\hat{p}_n} - f_p||_u$,

$$[1 - C||\tilde{\psi}_n^*||_2 \cdot (S(M_n) + T(M_n))] \cdot ||\hat{p}_n - p||_u \leq c \cdot [b_n^q w_q(b_n) + ||\tilde{\psi}_n^*||_2 \cdot a_n]. \tag{34}$$

Proposition 4.1 provides a general upper bound for $||\hat{p}_n - p||_u$ from $||f_{\hat{p}_n} - f_p||_u$, when $\tilde{h}$ has zeros. The parameters M_n, b_n, v_n^* are chosen such that the coefficient of $||\hat{p}_n - p||_u$ in (34) is bounded below by a positive constant and $||\tilde{\psi}_n^*||_2 \cdot a_n$ converges to zero. In the proof it is seen why $T(M_n)$ and $S(M_n)$ are defined using L_2-distance.

Example 4.1 *We evaluate (34) when h is either smooth or super-smooth, obtaining the same rates with the previous section. Since $\tilde{h}(t) \neq 0$, $\tilde{h}_n^* = \tilde{h}_n(t)$, and instead of $v_{n,j}$ we use*

$$v_n = \inf\{|\tilde{h}(t)|, ||t|| \leq M/b_n\} = \tilde{h}(M/b_n), \tag{35}$$

$$||\tilde{\psi}_n^*||_2 = [\int_{[-M/b_n, M/b_n]^d} \frac{|\tilde{K}(tb_n)|^2}{|\tilde{h}_n(t)|^2} dt]^{.5} \leq C b_n^{-.5d} v_n^{-1},$$

and (34) becomes

$$[1 - Cb_n^{-.5d} v_n^{-1} \cdot T(M_n)] \cdot ||\hat{p}_n - p||_u \leq b_n^q w_q(b_n) + C b_n^{-.5d} v_n^{-1} \cdot a_n \tag{36}$$

It is seen below that for selected M_n the term

$$C \cdot b_n^{-.5d} v_n^{-1} \cdot T(M_n) \tag{37}$$

converges to zero. Observe that the upper bound in (36) coincides with that in (15).

For the super-smooth model (17) and M_n large, from (35)

$$v_n^{-1} \leq C e^{\sum_{j=1}^d \alpha_j M^k b_n^{-k}} = C e^{d\bar{\alpha} M^k b_n^{-k}},$$

for $k \geq 1$,

$$T(M_n) \leq C[\Pi_{j=1}^d \int_{M_n}^{\infty} e^{-\alpha_j t_j^k} dt]^{.5} \leq C[\Pi_{j=1}^d \int_{M_n}^{\infty} e^{-\alpha_j t_j} dt]^{.5} \leq C e^{-.5 d\bar{\alpha} M_n},$$

for $0 < k < 1$,

$$T(M_n) \le c[\Pi_{j=1}^d \int_{M_n^k}^{\infty} e^{-\alpha_j y_j} y_j^{k^{-1}-1} dy_j]^{.5} \le c[\Pi_{j=1}^d \int_{M_n^k}^{\infty} e^{-.5\alpha_j y_j} dy_j]^{.5} \le ce^{-.25d\bar{\alpha}M_n^k}.$$

For the smooth model (18) and M_n *large, from (35)*

$$v_n^{-1} \sim b_n^{-d\bar{\beta}},$$

$$T(M_n) \le C[\Pi_{j=1}^d \int_{M_n}^{\infty} t^{-2\beta_j} dt]^{.5} \le cM_n^{.5\sum_{j=1}^d(-2\beta_j+1)} = cM_n^{-d(\bar{\beta}-.5)}.$$

Thus, for both models, (37) converges to zero when M_n*, i.e.* m*, is large enough such that (33) holds.*

The Oscillatory Model $(\mathcal{M}3)$

$(\mathcal{M}3)$ h is oscillatory decreasing at algebraic rate if

$$C_1|\sin(t)|^{\mu}|t|^{-\beta} \le |\tilde{h}(t)| \le C_2|\sin(t)|^{\mu}|t|^{-\beta}, |t| > T^* > 0, \tag{38}$$

and $\tilde{h}(t)$ does not vanish for $|t| \le T^*$; $\mu \ge 1,\ \beta > .5,\ 0 < C_1 \le C_2 < \infty$.

The *Oscillatory Model* is introduced in Hall and Meister (2007). The parameter μ describes the order of the isolated, periodic zeros of $\tilde{h}$; we assume μ is positive integer. Self-convolved uniform densities have Fourier transforms satisfying (38). Without loss of generality we use $\sin(t)$ instead of $\sin(\lambda t)$ and $\beta > .5$.

Before presenting the next proposition we study a motivating, special case.

Example 4.2 *Let*

$$\tilde{h}(t) = \frac{\sin t}{t}.$$

Assume that p *is* $\tilde{q}$*-smooth,* $\tilde{q} = q + \gamma$*. It is shown that in probability,*

$$||\hat{p}_n - p||_u \le Ca_n^{\tilde{q}/(\tilde{q}+3.5+\zeta)}, \tag{39}$$

for ζ *positive, but as close as we like to zero. Periodicity of* $\tilde{h}$*'s zeros implies that their number in the interval* $[-M_n, M_n]$ *is*

$$N(M_n) = c^* M_n.$$

The roots and parameters $v_{n,j}$ are indexed from the smallest to the largest, using positive indices for positive roots, $j = -.5c^*M_n, \ldots, .5c^*M_n$. Because of symmetry, the positive roots are used to provide upper bounds for $T, S, ||\tilde{\psi}_n^*||_2$. Let v_n be positive, decreasing to zero with n,

$$M_n = \frac{2M}{b_n^{2m}},\ m \geq .5, \qquad v_{n,j} = \frac{v_n}{j^{1+\delta}},\ j = 1, \ldots, .5c^*M_n;$$

δ to be determined for the best obtainable bound. For large $t > 0$, $|\tilde{h}(t)|^2$ is bounded by t^{-2}, thus

$$T(M_n) \leq CM_n^{-.5} = Cb_n^m.$$

Let L denote Lebesgue measure on the real line. To calculate $L(R_j)$ assume w.l.o.g. that $\tilde{h}$ is decreasing in a neighborhood of the j-th positive root, $r_j = j\pi$. Let

$$r_{j-} < r_j < r_{j+}:\ \tilde{h}(r_{j-}) = v_{n,j} = \frac{v_n}{j^{1+\delta}},\ \tilde{h}(r_{j+}) = -v_{n,j} = -\frac{v_n}{j^{1+\delta}}.$$

Make a first order Taylor expansion of $\tilde{h}(r_{j-})$ around root r_j of $\tilde{h}$:

$$\frac{v_n}{j^{1+\delta}} = \tilde{h}(r_{j-}) = (r_{j-} - r_j)\frac{\cos(y)y - \sin(y)}{y^2},\ r_{j-} < y < r_j = j\pi. \tag{40}$$

For large n, v_n is near zero and $r_{j-},\ r_{j+}$ near root r_j thus, from (40),

$$\frac{v_n}{j^\delta} \sim |r_{j-} - r_j| \sim |r_{j+} - r_j| \sim L(R_j), \tag{41}$$

$$S(M_n) = ||\tilde{h}_n - \tilde{h}_n^*||_2 \leq C[\sum_{j=-.5c^*M_n}^{.5c^*M_n} \int_{R_j} (\frac{v_n}{j^{1+\delta}} - |h(t)|)^2 dt]^{.5} \leq C[v_n^2 \sum_{j=1}^{.5c^*M_n} j^{-2(1+\delta)} v_n j^{-\delta}]^{.5}. \tag{42}$$

In the interval $[-M/b_n, M/b_n]$, $\tilde{h}$ has C^*M/b_n roots. By symmetry, we use only positive roots. By periodicity, intervals determined by successive roots have the same length and in the interval $J_j = [r_{j-1}, r_j]$ h_n^*'s smallest value is $\frac{v_n}{j^{1+\delta}}; r_0 = 0, j = 1, \ldots, .5C^*M/b_n$.

$$||\tilde{\psi}_n^*||_2 \leq C_1[\sum_{j=1}^{.5C^*M/b_n - 1} \int_{J_j} \frac{C_2}{|\tilde{h}_n^*(t)|^2} dt + \int_{J_{.5C^*M/b_n} \cup R_{.5C^*M/b_n}} \frac{C_2}{|\tilde{h}_n^*(t)|^2} dt]^{.5}$$

$$\leq C[\sum_{j=1}^{.5C^*M/b_n} j^{2(1+\delta)} v_n^{-2}]^{.5} \leq C[v_n^{-2} b_n^{-2(1+\delta)-1}]^{.5} \leq Cv_n^{-1} b_n^{-(1.5+\delta)}. \tag{43}$$

The best rate is obtained when $\delta = -1/3$ *and*

$$S(M_n) \leq Cv_n^{1.5}[-2m \ln b_n]^{.5}, \ ||\tilde{\psi}_n^*||_2 \leq cv_n^{-1}b_n^{-3.5/3}. \tag{44}$$

When $\delta \geq 0$ *a slower convergence rate is obtained,* 4.5 *replaces* 3.5 *in (39); with negative* δ*-values smaller than* $-1/3$ *one of the terms* $T(M_n)||\psi_n^*||_2$ *and* $S(M_n)||\psi_n^*||_2$ *increases to infinity with n. Replacing (44) in (34), with* $\delta = -1/3$*, we obtain*

$$[1 - Cv_n^{.5} \cdot b_n^{-3.5/3}[-2m \ln b_n]^{.5} - Cv_n^{-1}b_n^{m-3.5/3}] \cdot ||\hat{p}_n - p||_u \leq C[b_n^{\tilde{q}} + v_n^{-1}b_n^{-3.5/3}a_n]. \tag{45}$$

The value of m is determined such that the coefficient of $||\hat{p}_n - p||_u$ *in (45) is positive.*

Let

$$v_n = b_n^{2m/3}.$$

Then, the second and third terms in the coefficient of $||\hat{p}_n - p||_u$ *in (45) become, respectively,*

$$Cb_n^{(m-3.5)/3}[-2m \ln b_n]^{.5}, \qquad Cb_n^{(m-3.5)/3} \tag{46}$$

and taking

$$m = 3.5 + \xi, \ \xi > 0, \tag{47}$$

$||\hat{p}_n - p||_u$*'s coefficient is positive, smaller than one,* a_n*'s coefficient is*

$$b_n^{(-2m-3.5)/3} = b_n^{-3.5-2\xi/3}$$

and the rate of convergence is

$$b_n^{\tilde{q}} \sim a_n^{\tilde{q}/[\tilde{q}+3.5+2\xi/3]},$$

for any ξ *close to zero. Thus, (39) holds with* $\zeta = 2\xi/3$.

Remark 4.1 *In (39),* ζ *in the upper bound can be replaced by* $\zeta_n \downarrow 0$*. In (47) replace* ξ *by* $3\delta_n$ *and the terms in (46) become, respectively,*

$$Cb_n^{\delta_n}[-2m \ln b_n]^{.5}, \qquad Cb_n^{\delta_n}.$$

Choose δ_n :

$$b_n^{\delta_n} = \frac{1}{4C[-2m \ln b_n]^{.5}},$$

which implies that in (45) $||\hat{p}_n - p||_u$*'s coefficient is positive (since* b_n *will converge to zero),* a_n*'s coefficient is* $b_n^{-3.5-2\delta_n}$ *and* ζ_n *is* $2\delta_n$*. This remark holds also for the next proposition.*

Bound (34) is evaluated for the oscillatory model (38).

Proposition 4.2 *Under assumptions* $(\mathcal{A}1^*)$, $(\mathcal{A}2)-(\mathcal{A}5)$ *and the oscillatory model (38), in probability*

$$||\hat{p}_n - p||_u \leq C \cdot a_n^{\tilde{q}/(\tilde{q}+\beta+2\mu+.5+\zeta)}, \tag{48}$$

for any $\zeta > 0$.

Remark 4.2 *When* $\mu = 0$, *the upper bound in (48) becomes* $C \cdot a_n^{\tilde{q}/(\tilde{q}+\beta+.5+\zeta)}$, *for any* $\zeta > 0$, *i.e. misses bound (21) for smooth* h *when* d *is unity by a factor increasing slower than any power of* n. *The denominator in the exponent of* a_n *makes the bound similar to that in Meister (2008, Theorem 1) which is function of all the parameters in the problem but is obtained by combining moment and smoothness assumptions. In Hall and Meister (2007, Proposition 4.2), under additional assumptions, the upper bound on the mean square error of* $\hat{p}_n$ *is surprisingly function either of* $\tilde{q}$ *and* β *only or of* μ *only, i.e., these bounds are, respectively,* $n^{-1/2\mu}$, $n^{-1/2\mu}\log n$, $n^{-2\tilde{q}/(2\tilde{q}+2\beta+1)}$.

The upper bound in (48) could be improved , e.g., by using in the definition of $\tilde{h}_n^*$ *in (29)* $v_{n,j}$*-values obtained via ridging and also the bounds for* $V_{1,n}$ *and additional assumptions in Hall and Meister (2007, proof of Proposition 2) or in Meister (2008). We decided not to do so for having rates of convergence with the original assumptions, especially since* $\hat{p}_n$ *would be usually obtained from optimal estimate* $f_{\hat{p}_n}$. *We did not obtain bounds for the oscillatory super-smooth model due to the results in section 3, especially Corollary 3.2, that makes this case uninteresting.*

Example 4.3 *Assume that* p *is* q*-smooth, defined on a compact interval in* R. *Let* h *be the uniform density on* $[0, 1]$. *Then* f_p *is* q*-smooth in a compact interval in* R *and an* L_1*-optimal minimum distance estimate* $f_{\hat{p}_n}$ *of* f_p *can be obtained, e.g., using Yatracos (1985). The rate of convergence of* $f_{\hat{p}_n}$ *to* f_p *is* $n^{-q/(2q+1)}$ *and from Example 4.2 the* L_1*-upper error bound of* $\hat{p}_n$ *to* p *is* $[n^{-q/(2q+1)}]^{q/(q+3.5+\zeta)}$, *for* ζ *any positive number near zero.*

5 Appendix

Lemma 5.1 *Let g be a function defined on a set $\mathcal{C}$ in R^d that has all s-th mixed order partial derivatives for $0 \leq [s] \leq q$, with the q-th derivative having modulus of continuity w_q. Then, for kernel K satisfying (9) and K_n defined in (10):*
a) If C is compact,

$$||g - K_n * g||_u \leq C b_n^q w_q(b_n), \ u \geq 1. \tag{49}$$

b) If C is R^d and g is bounded, then (49) holds for the L_∞-distance.

Proof of Lemma 5.1: *a)* The result follows from Yatracos (1989,p. 173, Proposition 1). *b)* Follows from Shapiro (1969, p. 52, Theorem 20) when $d = 1$, and for $d > 1$ from Yatracos (1989) because the bound for $|g(x) - K_n * g(x)|$ remains valid for any $x \in R^d$ and in L_∞. $\square$

Proof of Lemma 3.1: For the Fourier transform $\tilde{K}_n(x)$ it holds,

$$\tilde{K}_n(x) = C \int e^{-ity} b_n^{-d} K(y/b_n) dy = C \int e^{i(tb_n)yb_n^{-1}} K(yb_n^{-1}) d(yb_n^{-1}) = C\tilde{K}(xb_n).$$

Boundedness of $\tilde{K}$, the Cauchy-Schwartz inequality and Parseval's identity imply that

$$||\psi_n||_1 \leq [\int |\psi_n(x)|^2 dx]^{.5} = C[\int |\tilde{\psi}_n(t)|^2 dt]^{.5} = C[\int \frac{|\tilde{K}_n(t)|^2}{|\tilde{h}(t)|^2} dt)]^{.5}$$

$$= C[\int_{[-M/b_n, M/b_n]^d} \frac{|\tilde{K}(b_n t)|^2}{|\tilde{h}(t)|^2} dt]^{.5} \leq C \cdot [\int_{[-\frac{M}{b_n}, \frac{M}{b_n}]^d} |\tilde{h}(t)|^{-2} dt \ ]^{1/2}$$

$$\leq C \frac{\sup_{t \in [-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1}}{b_n^{.5d}}. \qquad \square$$

Proof of Proposition 3.1: *a)* For $1 \leq u < \infty$, it holds:

$$[\int_{\mathcal{Y}} |\hat{p}_n(y) - p(y)|^u dy]^{1/u} \leq [\int_{\mathcal{Y}} |\hat{p}_n(y) - K_n * \hat{p}_n(y)|^u dy]^{1/u}$$

$$+[\int_{\mathcal{X}} |K_n * \hat{p}_n(x) - K_n * p(x)|^u dx]^{1/u} + [\int_{\mathcal{Y}} |K_n * p(y) - p(y)|^u dy]^{1/u}$$

$$\leq C b_n^q w_q(b_n) + ||\psi_n * h * (\hat{p}_n - p)||_u \leq C b_n^q w_q(b_n) + ||\psi_n||_1 \cdot ||f_{\hat{p}_n} - f_p||_u.$$

The first inequality is due to the triangular property of the $||\cdot||_u$-distance and to $\mathcal{Y} \subset \mathcal{X}$. The second inequality is due to Lemma 5.1 and (13). The third inequality follows from Young's inequality for convolutions. The result follows from Lemma 3.1.

For the sup-norm L_∞- bound observe that

$$|\hat{p}_n(y) - p(y)| \leq |\hat{p}_n(y) - K_n * \hat{p}_n(y)| + |K_n * \hat{p}_n(y) - K_n * p(y)| + |K_n * p(y) - p(y)|$$

$$\leq Cb_n^q w_q(b_n) + |\psi_n * h * (\hat{p}_n - p)(y)| \leq Cb_n^q w_q(b_n) + \int |\psi_n(v) h * (\hat{p}_n - p)(y - v)| dv$$

$$\leq Cb_n^q w_q(b_n) + ||\psi_n||_1 \cdot ||h * p_n - h * p||_\infty.$$

The last inequality is obtained by bounding $|h * (p_n - p)(y - v)|$ in the integral with its supremum over all v.

b) When p is bounded and has domain R^d, the upper bound in L_∞ is obtained as above.
□

Proof of Proposition 3.2: *a) i)* When h follows the super-smooth model (17), the second term in the upper bound (15) has an exponential rate but the first term decreases at algebraic rate. Since

$$\sup_{t \in [-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1} \leq C \cdot e^{\sum_{j=1}^d \alpha_j M^k b_n^{-k}} \leq C \cdot e^{d\bar{\alpha} M^k b_n^{-k}}, \tag{50}$$

the second term in upper bound (15) converges to zero as n increases if

$$\lim_{n \to \infty} \frac{\exp\{d\bar{\alpha} M^k b_n^{-k}\}}{b_n^{.5d}} a_n = 0 \leftrightarrow \lim_{n \to \infty} d\bar{\alpha} M^k b_n^{-k} - .5d \log b_n - \log a_n^{-1} = -\infty. \tag{51}$$

Choosing

$$b_n^k = \frac{4d\bar{\alpha} M^k}{\log a_n^{-1}} \text{ or } b_n = \frac{(4d\bar{\alpha})^{1/k} M}{(\log a_n^{-1})^{1/k}}$$

(51) holds and the terms in upper bound (15) are

$$b_n^q w_q(b_n) \sim (\log a_n^{-1})^{-q/k} w_q[C(\log a_n^{-1})^{-1/k}], \tag{52}$$

$$\frac{\sup_{t \in [-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1}}{b_n^{.5d}} a_n \leq a_n^{3/4} (\log a_n^{-1})^{5d/k}, \tag{53}$$

with (53) converging faster to 0 as n increases than (52).

When $w_q(b_n) \sim b_n^{\gamma}$, (52) determines the upper convergence rate $(\log a_n^{-1})^{-(q+\gamma)/k}$.

ii) When h follows the smooth model (18), both terms in upper bound (15) have algebraic rate. Since

$$\sup_{t \in [-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1} \leq C \cdot (\frac{M}{b_n})^{d\bar{\beta}}$$

we choose b_n such that

$$b_n^q w_q(b_n) \sim \frac{a_n}{b_n^{d\bar{\beta}+.5d}}.$$

When $w_q(b_n) \sim b_n^{\gamma}$, $\tilde{q} = q + \gamma$,

$$b_n^{\tilde{q}} \sim \frac{1}{b_n^{d\bar{\beta}} \cdot b_n^{.5d}} a_n \text{ or } b_n \sim a_n^{1/(\tilde{q}+d\bar{\beta}+.5d)} \tag{54}$$

and

$$||\hat{p}_n - p||_u \leq c_M a_n^{\tilde{q}/(\tilde{q}+d\bar{\beta}+.5d)}.$$

iii) Follows using the approach in *i)* and *ii).*

b) The results follow as in *i)-iii).* $\square$

Proof of Corollary 3.1: *a)* Follows along the lines in Yatracos (1989), Proposition 2, p. 174 and Remarks (i) and (ii) pages 174, 175, since p and $p^{(s)}$ vanish outside their domain.

b) When p is defined in R^d, the results still hold since in the $u - v$ integration by parts which allows to pass from $p^{(s)}$ a derivative to the kernel, the $u \cdot v$ term vanishes at infinities. $\square$

Proof of Corollary 3.2: The bounds are obtained by plugging $a_n \sim n^{-\delta}$ in the bounds in Proposition 3.2 *a) i)* and in (22) and (23). For densities in R, optimality for any $\delta > 0$ follows from the optimal rates in Fan (1991,1992, 1993). $\square$

Proof of Proposition 4.1: Along the lines of proof for Proposition 3.1. Bounds are provided for the difference of convolutions of K_n with $\hat{p}_n$ and p. Using the triangular inequality, (32), properties of Fourier transforms and repeatedly that for g_1, $\tilde{g}_1$ and g_2,

$$||g_1 * g_2||_u \leq ||g_1||_1 \cdot ||g_2||_u \leq c \cdot ||\tilde{g}_1||_2 \cdot ||g_2||_u,$$

we have

$$||K_n * (\hat{p}_n - p)||_u = ||\psi_n^* * h_n^* * (\hat{p}_n - p)||_u = ||\psi_n^* * (h_n^* - h_n + h_n - h + h) * (\hat{p}_n - p)||_u$$

$$\leq ||\psi_n^* * (h_n^* - h_n) * (\hat{p}_n - p)||_u + ||\psi_n^* * (h_n - h) * (\hat{p}_n - p)||_u + ||\psi_n^* * h * (\hat{p}_n - p)||_u$$

$$\leq ||\psi_n^*||_1 \cdot ||(h_n^* - h_n) * (\hat{p}_n - p)||_u + ||\psi_n^*||_1 \cdot ||(h_n - h) * (\hat{p}_n - p)||_u + ||\psi_n^*||_1 \cdot a_n$$

$$\leq c[||\psi_n^*||_1 \cdot ||\tilde{h}_n - \tilde{h}_n^*||_2 \cdot ||\hat{p}_n - p||_u + ||\psi_n^*||_1 \cdot ||\tilde{h} - \tilde{h}_n||_2 \cdot ||\hat{p}_n - p||_u + ||\psi_n^*||_1 \cdot a_n].$$

The result follows from (27), (30) and

$$||\psi_n^*||_1 \leq C \cdot ||\tilde{\psi}_n^*||_2. \qquad \square$$

Proof of Proposition 4.2: The same steps are followed as in Example 4.2 with only change the $v_{n,j}$-values. Let

$$M_n = \frac{2M}{b_n^{2m}},\ m \geq .5, \qquad v_{n,j} = \frac{v_n}{j^{\beta+\delta}},\ \delta \geq 0,\ j = 1, \ldots, c^* M_n.$$

For $T(M_n)$ it holds,

$$T(M_n) \leq C[\int_{M_n}^{\infty} \frac{1}{t^{2\beta}} dt]^{.5} = C[\frac{t^{-2\beta+1}}{-2\beta+1}|_{M_n}^{\infty}]^{.5} \sim M_n^{-(\beta-.5)} \sim b_n^{m(2\beta-1)}.$$

To calculate $L(R_j)$ for (38) make a Taylor expansion of $\tilde{h}(r_{j-})$ around r_j. The first non-zero coefficient is that of $(r_{j-} - r_j)^\mu / j^\beta$ which implies that

$$L(R_j) \sim (\frac{v_n}{j^\delta})^{1/\mu}.$$

Then,

$$S(M_n) \leq C[v_n^2 \sum_{j=1}^{.5c^* M_n} j^{-2(\beta+\delta)} L(R_j)]^{.5} \sim v_n^{1+\frac{1}{2\mu}} [\sum_{j=1}^{.5c^* M_n} j^{-2\beta - \delta(\frac{2\mu+1}{\mu})}]^{.5}, \tag{55}$$

$$||\tilde{\psi}_n^*||_2 \leq C_1[\sum_{j=1}^{.5C^* M/b_n - 1} \int_{J_j} \frac{C_2}{\tilde{h}_n^*(t)^2} dt + \int_{J_{.5C^* M/b_n} \cup R_{.5C^* M/b_n}} \frac{C_2}{\tilde{h}_n^*(t)^2} dt]^{.5}$$

$$\leq C[\sum_{j=1}^{.5C^* M/b_n} j^{2(\beta+\delta)} v_n^{-2}]^{.5} \leq C[v_n^{-2} b_n^{-2(\beta+\delta)-1}]^{.5} \leq C v_n^{-1} b_n^{-(\beta+\delta+.5)}. \tag{56}$$

Replacing in (55), (56),

$$\delta = \frac{\mu(1-2\beta)}{2\mu+1}, \tag{57}$$

the corresponding bounds for $S(M_n)$ and $||\tilde{\psi}_n^*||_2$ become, respectively,

$$Cv_n^{1+\frac{1}{2\mu}}[-2m\ln b_n]^{.5}, \quad Cv_n^{-1}b_n^{-\frac{\beta+2\mu+.5}{2\mu+1}}. \tag{58}$$

Thus, (34) becomes

$$\{1-Cb_n^{-\frac{\beta+2\mu+.5}{2\mu+1}}[v_n^{1/2\mu}(-2m\ln b_n)^{.5}-v_n^{-1}b_n^{m(2\beta-1)}]\}\cdot||\hat{p}_n-p||_u \leq C[b_n^q w_q(b_n)+v_n^{-1}b_n^{-\frac{\beta+2\mu+.5}{2\mu+1}}\cdot a_n]. \tag{59}$$

Let

$$v_n = b_n^{\frac{2\mu}{2\mu+1}m(2\beta-1)}.$$

Then, the second and third terms in the coefficient of $||\hat{p}_n-p||_u$ in (59) become, respectively,

$$Cb_n^{\frac{1}{2\mu+1}[m(2\beta-1)-(\beta+2\mu+.5)]}[-2m\ln b_n]^{.5}, \quad Cb_n^{-\frac{1}{2\mu+1}[m(2\beta-1)-(\beta+2\mu+.5)]}$$

and taking

$$m(2\beta-1) = \beta+2\mu+.5+\xi, \ \xi > 0,$$

$||\hat{p}_n-p||_u$'s coefficient is positive, smaller than one, a_n's coefficient is

$$b_n^{-(\beta+2\mu+.5)-\frac{2\mu\xi}{2\mu+1}}$$

and the rate of convergence is

$$b_n^{\tilde{q}} \sim a_n^{\tilde{q}/[\tilde{q}+\beta+2\mu+.5+(2\mu\xi)/(2\mu+1)]}.$$

Thus, (39) holds with $\zeta = 2\mu\xi/(2\mu+1)$. $\Box$

References

[1] Carrol, R. J. and Hall, P. (1988) Optimal rates of convergence for deconvolving a density. *JASA* **83**, 1184-1186.

[2] Comte, F. and Lacour, C. (2013) Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.* **49**, 569-609.

[3] Delaigle, A. and Gijbels, I. (2002). Estimation of integrated squared density derivatives from a contaminated sample. *JRSS B* **64**, 869-886.

[4] Devroye, L. (2013) Personal Communication

[5] Devroye, L. (1992) A note on the usefulness of superkernels in density estimation. *Ann. Statist.* **20**, 2037-2056.

[6] Devroye, L. (1989) Consistent deconvolution in density estimation.*Canad. J. Statist* **17**, 235-239.

[7] Fan, J. (1993) Adaptive local one-dimensional sub-problems with application to a deconvolution problem. *Ann. Statist.* **21**, 600-610.

[8] Fan, J. (1992) Deconvolution with supersmooth distributions. *Can. J. Stat.* **20**, **2**, 155-169.

[9] Fan, J. (1991) On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.

[10] Genovese, C. R. and Wasserman, L. (2000) Rates of convergence for the Gaussian mixture sieve. *Ann. Statist.* **28**, 1105-1127.

[11] Ghosal, S. and van der Vaart, A. W. (2001) Entropy and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29**, 1233-1263.

[12] Groeneboom, P. and Jongbloed, G. (2003) Density estimation in the uniform deconvolution model. *Statist. Nederlandica* **57**, 136-157.

[13] Hall, P. and Meister, A. (2007) A ridge-parameter approach to deconvolution. *Ann. Stat.* **35**, 1535-1558.

[14] Hesse, C. H. (1995) Deconvolving a density from partially contaminated observations. *J. Multivariate Anal.* **55**, 246-260.

[15] Ibragimov, I. A. (2001) Estimation of analytic functions. *State of the Art in Probability and Statistics. Festschrift for W. R. van Zwet.* IMS, Hayward.

[16] Johaness, J. (2009) Deconvolution with unknown error distribution. *Ann. Stat.* **37**, 2301-2323.

[17] Meister, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems* **24**, **1**, 015003.

[18] Meister, A. (2006). Density estimation with normal measurement error with unknown variance. Statist. Sinica **16**, 195-211.

[19] Loh, W- L. and Zhang, C.-H. (1996) Global properties of kernel estimators for mixing densities in exponential family models for discrete variables. *Statist. Sinica* **6**, 561-678.

[20] Loh, W- L. and Zhang, C.-H. (1997) Estimating mixing densities in exponential family models for discrete variables. *Scand. J. Statist.* **24**, 15-32.

[21] Rebelles, G. (2015) Structural adaptive deconvolution under L_p-losses. *arXiv:1504.06246v2*

[22] Robbins, H. (1964) The empirical Bayes approach to statistical decision problems. *Ann. Math. Stat.* **35**, 1-20.

[23] Robbins, H. (1955) An empirical Bayes approach to statistics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* p. 131-148.

[24] Robbins, H. (1948) Mixture of Distributions. *Ann. Math. Stat.* **19**, 3, 360-369.

[25] Shapiro (1969) *Smoothing and approximation of functions.* Van Nostrand Reinhold Company, New York.

[26] Stefanski, L. A. and Carroll, R. J. (1990) Deconvoluting kernel density estimators. *Statistics* **21**, 169-184.

[27] Yatracos, Y. G. (1989) On the estimation of the derivatives of a function with the derivatives of an estimate. *J. of Multivariate Analysis* **28**, 172-175.

[28] Yatracos, Y. G. (1985) Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *Ann. Statist.* **13**, 768-774.

[29] Zhang, C.-H. (2009) Generalized maximum likelihood estimation of normal mixture densities. *Statist. Sinica* **19**, 1297-1318.

[30] Zhang, C.-H. (1990) Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806-831.